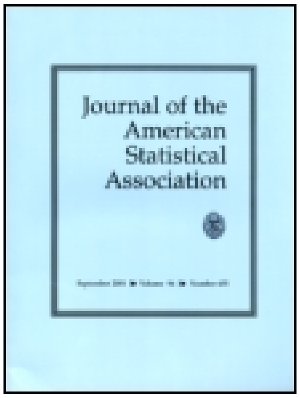

Journal of the American Statistical Association
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/uasa20

A Smooth Simultaneous Confidence Corridor for the Mean of Sparse Functional Data

Shuzhuan Zheng, Lijian Yang & Wolfgang K. Härdle

Accepted author version posted online: 03 Dec 2013.Published online: 13 Jun 2014.

To cite this article: Shuzhuan Zheng, Lijian Yang & Wolfgang K. Härdle (2014) A Smooth Simultaneous Confidence Corridor for the Mean of Sparse Functional Data, Journal of the American Statistical Association, 109:506, 661-673, DOI: 10.1080/01621459.2013.866899

To link to this article: http://dx.doi.org/10.1080/01621459.2013.866899

A Smooth Simultaneous Confidence Corridor for the Mean of Sparse Functional Data

Shuzhuan ZHENG, Lijian YANG, and Wolfgang K. HÄRDLE

Functional data analysis (FDA) has become an important area of statistics research in the recent decade, yet a smooth simultaneous confidence corridor (SCC) does not exist in the literature for the mean function of sparse functional data. SCC is a powerful tool for making statistical inference on an entire unknown function, nonetheless classic "Hungarian embedding" techniques for establishing asymptotic correctness of SCC completely fail for sparse functional data. We propose a local linear SCC and a shoal of confidence intervals (SCI) for the mean function of sparse functional data, and establish that it is asymptotically equivalent to the SCC of independent regression data, using new results from Gaussian process extreme value theory. The SCC procedure is examined in simulations for its superior theoretical accuracy and performance, and used to analyze growth curve data, confirming findings with quantified high significance levels. Supplementary materials for this article are available online.

KEY WORDS: Double sum; Extreme value; Karhunen–Loève L^2 representation; Local linear estimator; Strong approximation.

1. INTRODUCTION

Functional data analysis (FDA) has become an important area of statistics research in the recent decade, see, for instance, Cardot, Ferraty, and Sarda (2003) and Cai and Hall (2006). One well-known application is growth curve analysis in biology, medicine, and chemistry, see, for example, Müller (2009), James, Hastie, and Sugar (2000), and Ferraty and Vieu (2006), and references therein. Much of the existing work though is devoted to consistency of estimation and/or dimension reduction. Results on statistical inference for the mean curve are rather scarce although it is important for characterization of important data features. To characterize global properties of the unknown function of interest, a simultaneous confidence corridor (SCC), or a shoal of confidence intervals (SCI), is the appropriate instruments.

An SCC is a collection of sliding random intervals placed above each point of the data range, so that the scanned two-dimensional random region in the shape of a corridor (or band) contains an entire unknown curve (or function) with predetermined probability. An SCI, on the other hand, consists of sliding random intervals containing each point of the unknown curve with predetermined probability. Although the SCI is also a two-dimensional random region as SCC, it does not contain the entire curve with a positive probability. Nonetheless, SCI is widely used in the literature because it is usually straightforward to compute and illustrates well the overall trend and shape of the unknown curve, see Figure 6 of Xue and Yang (2006) and Figure 1 of Wang and Yang (2007).

Decisions about the mean curve of functional data are critical, for example, in ozone analysis, see Lucas and Diggle (1997) for a longitudinal study on Sitka spruce. They pointed out that, to assess the cumulative effect of ozone pollution on spruce, an inference on the mean function of spruce growth during the entire experiment rather than at the end of the growth is required. This is one of the many motivations to develop a new method to construct an SCC for the mean function of sparse functional data, where the measurements are randomly located with random repetitions.

To illustrate the use of SCC and SCI, consider the growth curve data in the study of human skeletal health, consisting of measurements Y_{ij}, the spinal bone mineral density (g/cm^2), for $n = 132$ subjects (nonblack females). The number of measurements for the ith subject, N_i, varies randomly between 2 and 4 (sparsity), and X_{ij}, the jth time point of measurement for the ith subject (aged 8.8–26.2 yr), also varies randomly for each subject. James, Hastie, and Sugar (2000) used bootstrap confidence intervals to test the mean function of these data at points of interest, for example, the growth peak at about 15 yr.

Figure 2(a) exhibits the scatterplot of the female spinal bone density versus the age. Figure 2(b) depicts the SCCs and SCIs of the population mean growth curve at the confidence level of 90%. Our SCIs are similar to those of James, Hastie, and Sugar (2000) for pointwise inference, while our SCCs indicate that the mean spinal bone density level increases with age, but the bone growth is accelerated during early adolescence (9–15 yr), whereas it reaches the plateau during late puberty (16–26 yr).

The SCC construction has been extensively studied in the literature. Neumann and Polzehl (1998), Kreiss and Neumann (1998), and Claeskens and Van Keilegom (2003) developed

Shuzhuan Zheng is Visiting Scholar, Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou 215006, China, and Ph.D. student, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824 (E-mail: zheng@stt.msu.edu). Lijian Yang is Director, Center for Advanced Statistics and Econometrics Research, Soochow University, Suzhou 215006, China, and Professor, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824 (E-mail: yanglijian@suda.edu.cn; yang@stt.msu.edu). Wolfgang K. Härdle is Professor, C.A.S.E.-Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany, and Distinguished Visiting Professor, Lee Kong Chian School of Business, Singapore Management University (E-mail: haerdle@wiwi.hu-berlin.de). This work is supported in part by the Deutsche Forschungsgemeinschaft through the CRC 649 "Economic Risk," NSF Awards DMS 0706518, DMS 1007594, an MSU Dissertation Continuation Fellowship, funding from the National University of Singapore, the Jiangsu Specially Appointed Professor Program SR10700111 and the Jiangsu Key Discipline Program (Statistics) ZY107002, Jiangsu, China. Insightful comments from the Editor, the Associate Editor, and three referees have led to significant improvement of both the content and presentation of the article.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/jasa.

Journal of the American Statistical Association
June 2014, Vol. 109, No. 506, Theory and Methods
DOI: 10.1080/01621459.2013.866899

bootstrap methods, Kerkyacharian, Nickl, and Picard (2012) established a nonasymptotic SCC via concentration inequalities, while another strand of literature including Bickel and Rosenblatt (1973), Johnston (1982), Xia (1998), and Fan and Zhang (2000) constructed the SCC by Gaussian strong approximation and extreme value theory, which though did not assume that for family of curves one needs to take care of dependence structures. By contrast, Wu and Zhao (2007) constructed a confidence band for the nonstationary mean function, and Wang and Yang (2007), and Song and Yang (2009) obtained the spline-based analogy for the mean and variance functions. Nonparametric time series with specific dependence structures were considered in Zhao and Wu (2008). For sparse functional data, Yao, Müller, and Wang (2005a) constructed prediction bands for individual trajectories instead of confidence band for the mean function, while Yao (2007) obtained an SCI for the mean and covariance functions. More recently, asymptotic SCC based on spline regression has been constructed for the mean function of sparse functional data in Ma, Yang, and Carroll (2012), of dense functional data in Cao, Yang, and Todem (2012), and for mean function derivatives of dense functional data in Cao et al. (2012).

The main difficulty in constructing SCCs for functional data is that the observations within subject are dependent, thus the "Hungarian embedding," widely used in the existing literature to construct nonparametric confidence bands (e.g., Johnston 1982; Claeskens and Van Keilegom 2003), is no longer available. Ma, Yang, and Carroll (2012) made use of partial sum strong approximation to reduce the error process in estimating mean function to a Gaussian sequence whose maximal deviation asymptotics is obtained via Gaussian sequence extreme value theory. As the only published work that rigorously establishes SCC for sparse functional data, it suffers three serious flaws: it uses piecewise constant spline and thus consists of discontinuous step functions; it has width of order $n^{-1/3}\log n$, significantly wider than the typical order of $n^{-2/5}\log n$ for regression data, see, for example, Xia (1998); its SCC is valid only for processes with finite Karhunen–Loève L^2 representation.

We propose to construct the SCC for the mean function of sparse functional data via local linear smoothing, with more appealing theoretical and computational properties. Our local linear SCC has width of order $n^{-2/5}\log n$ instead of $n^{-1/3}\log n$, which is by no means insignificant. The local linear SCC and the estimated mean curve are also smooth, rather than step functions. Another major theoretical advance of the local linear SCC is that it is based on infinite number of positive eigenvalues of the covariance function, while Ma, Yang, and Carroll (2012) required the covariance function to have a finite number of positive eigenvalues. Reflective of these theoretical advantages, the local linear SCC clearly outperforms Ma, Yang, and Carroll (2012) in simulation: it is much narrower and thus highlights much more sharply the features of the mean function, with coverage frequencies much closer to nominal levels. The local linear SCC thus allows for much more informative and reliable global inference.

In addition to the superior inference features, proving asymptotics for the local linear SCC has added powerful technical tools to the arsenal of nonparametric smoothing. Instead of the Normal Comparison Lemma (also known as the Slepian Lemma in extreme value theory) for arrays of Gaussian random variables used in Ma, Yang, and Carroll (2012), we combine the highly complicated Kac–Rice Formula for exceedance probability of Gaussian process (Cierco-Ayrolles, Croquette, and Delmas 2003), and the Double Sum Method of Piterbarg (1996) to show that the maximal deviation of the local linear estimator for sparse functional mean converges to the same Gumbel distribution as if the data were actually iid. We are unaware of other works on SCC that employs these new techniques.

We organize our article as follows. In Section 2, we describe the model and local linear smoothing methodology. In Section 3, we present the asymptotic distribution of the maximal deviation of the local linear estimator from the true mean function, which is used to construct the SCC. Section 4 outlines the key steps to implement the SCC. Section 5 illustrates the performance of the SCC through extensive simulations followed by an empirical example in Section 7 which illustrates the SCC application on growth curve data. Technical proofs are presented in the Appendix.

2. MODEL AND METHODOLOGY

Longitudinal data have the form of $\{X_{ij}, Y_{ij}\}$, $1 \le j \le N_i$, $1 \le i \le n$, in which $X_{ij} \in \mathcal{X} = [a, b]$ is the jth random time point for the ith subject and Y_{ij} is the response measured at X_{ij}. For the ith subject, the sample path is the noisy realization of a continuous time stochastic process $\xi_i(x)$, namely,

$$Y_{ij} = \xi_i(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}, \tag{2.1}$$

where the errors ε_{ij} are iid with $\mathrm{E}\,\varepsilon_{ij} = 0, \mathrm{E}\,\varepsilon_{ij}^2 = 1$, and $\{\xi_i(x), x \in \mathcal{X}\}$ are iid copies of the process $\{\xi(x), x \in \mathcal{X}\}$ with $\mathrm{E}\int_{\mathcal{X}} \xi^2(x)dx < +\infty$.

Denote by $m(x) = \mathrm{E}\xi(x)$ the regression curve and by $G(x, x') = \mathrm{cov}\{\xi(x), \xi(x')\}$ the within-subject covariance function with the Karhunen–Loève L^2 representation

$$\xi_i(x) = m(x) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(x), \tag{2.2}$$

one has the random coefficients $\{\xi_{ik}\}_{k=1}^{\infty}$ uncorrelated with mean 0 and variance 1. Here, $\phi_k(x) = \sqrt{\lambda_k}\psi_k(x)$, where $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\psi_k(x)\}_{k=1}^{\infty}$ are respectively the eigenvalues and eigenfunctions of $G(x, x')$ such that $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ and $\{\psi_k\}_{k=1}^{\infty}$ forms an orthonormal basis of $L^2(\mathcal{X})$. Therefore, $G(x, x') = \sum_{k=1}^{\infty} \phi_k(x)\phi_k(x')$ and $\int G(x, x')\phi_k(x')dx' = \lambda_k\phi_k(x)$. Equations (2.1) and (2.2) can then be written as

$$Y_{ij} = m(X_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(X_{ij}) + \sigma(X_{ij})\varepsilon_{ij}. \tag{2.3}$$

For convenience, we denote the conditional variance of Y_{ij} given $X_{ij} = x$ as

$$\sigma_Y^2(x) = G(x, x) + \sigma^2(x) = \mathrm{var}(Y_{ij}|X_{ij} = x). \tag{2.4}$$

We are interested in the sparse situation where the number of measurements N_i within subject are iid copies of a positive random integer N_1, see Yao, Müller, and Wang (2005a, b) and Yao (2007).

To introduce the estimator, denote by K a kernel function, $h = h_n > 0$ a bandwidth and $K_h(x) = h^{-1}K(x/h)$.

Let $N_{\mathrm{T}} = \sum_{i=1}^{n} N_i$ be the total sample size and define $\mathbf{Y} = (Y_{ij})_{1\le j\le N_i, 1\le i\le n}$ the $N_{\mathrm{T}}\times 1$ vector of responses. For any $x \in [0,1]$, let $\mathbf{X} = \mathbf{X}(x) = (1, X_{ij} - x)_{1\le j\le N_i,1\le i\le n}$ be the design matrix for linear regression and $\mathbf{W} = \mathbf{W}(x) = N_{\mathrm{T}}^{-1}\operatorname{diag}\{K_h(X_{11}-x),\ldots,K_h(X_{nN_n}-x)\}$ the kernel weight diagonal matrix. Following Fan and Gijbels (1996), local linear estimators of $m(x)$ and $m'(x)$ are

$$\{\widehat{m}(x), \widehat{m}'(x)\}^{\mathrm{T}} = \arg\min_{a,b}\{\mathbf{Y} - \mathbf{X}(a,b)^{\mathrm{T}}\}^{\mathrm{T}}\mathbf{W}\{\mathbf{Y} - \mathbf{X}(a,b)^{\mathrm{T}}\} = (\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Y}.$$

Consequently, with $e_0^{\mathrm{T}} = (1,0)$, $\widehat{m}(x)$ is written as

$$\widehat{m}(x) = e_0^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{Y}, \tag{2.5}$$

where the dispersion matrix

$$\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X} = \operatorname{diag}(1,h)\begin{pmatrix} s_{n,0} & s_{n,1} \\ s_{n,1} & s_{n,2}\end{pmatrix}\operatorname{diag}(1,h), \tag{2.6}$$

and for any nonnegative integer l,

$$s_{n,l} = s_{n,l}(x) = N_{\mathrm{T}}^{-1}\sum_{i,j} K_h(X_{ij}-x)\{(X_{ij}-x)/h\}^l. \tag{2.7}$$

The local linear estimator $\widehat{m}(x)$ is preferred over other kernel type estimators because it uses local information of the data and is minimax efficient. It is also adaptive to the design and automatically corrects for boundary effects, see Fan and Gijbels (1996) for details.

3. MAIN RESULTS

Without loss of generality, assume $\mathcal{X} = [0,1]$ and consider the assumptions:

- (A1) The mean function $m(x) \in C^2[0,1]$, that is, it is twice continuously differentiable.
- (A2) $\{X_{ij}\}_{i=1,j=1}^{\infty,\infty}$ are iid with probability density function $f(x)$. The functions $f(x)$, $\sigma(x)$ and $\phi_k \in C^1[0,1]$ with $f(x) \in [c_f, C_f]$, $\sigma(x) \in [c_\sigma, C_\sigma]$ and all involved constants are finite and positive. The eigenfunction series is uniformly absolute convergent: $\sum_{k=1}^{\infty}\|\phi_k\|_\infty < C_\phi < +\infty$.
- (A3) The numbers of observations $N_i, i = 1,2,\ldots$ are iid random positive integers with $\mathrm{E}N_1^r \le r!c_N^r, r = 2,3,\ldots$ for some constant $c_N > 0$. $(N_i)_{i=1}^{\infty}$, $(X_{ij})_{i=1,j=1}^{\infty,\infty}$, $(\xi_{ik})_{i=1,k=1}^{\infty,\infty}$, $(\varepsilon_{ij})_{i=1,j=1}^{\infty,\infty}$ are independent, while $\{\xi_{ik}\}_{i=1,k=1}^{\infty,\infty}$ are iid N(0, 1).
- (A4) There exists $r > 5$, such that $\mathrm{E}|\varepsilon_{11}|^r < \infty$.
- (A5) The bandwidth $h = h_n$ satisfies $nh^4 \to \infty$, $nh^5\log n \to 0$ and $h < 1/2$.
- (A6) The kernel function $K(x)$ is a symmetric probability density function supported on $[-1,1]$ and $\in C^3[-1,1]$.

Assumptions (A1), (A2), (A5), and (A6) appear in many papers on kernel smoothing. Except normality of the ξ_{ik}'s, Assumption (A3) has been used in Yao, Müller, and Wang (2005a) and Paul and Peng (2009), while Assumption (A4) is in Ma, Yang, and Carroll (2012). Note that despite the independence Assumption (A3), temporal dependence in time course data such as growth curve data is nicely captured by model (2.3), as responses of the same subject i at different times are correlated: $\operatorname{cov}(Y_{ij}, Y_{ij'}|X_{ij}, X_{ij'}) = G(X_{ij}, X_{ij'})$, according to Equation (2.3), the definition of the within-subject covariance function G and Assumption (A3).

Assumptions (A2) and (A3) entail that $\mathrm{E}\sum_{k=1}^{\infty}\|\phi_k\|_\infty|\xi_{ik}| < +\infty$, thus the random variable $\sum_{k=1}^{\infty}\|\phi_k\|_\infty|\xi_{ik}| < +\infty$ almost surely. Therefore over an event of probability measure 1 and $x \in [0,1]$, $\sum_{k=1}^{\infty}\xi_{ik}\phi_k(x)$ converges uniformly. Since $\{\xi_{ik}\}_{i=1,k=1}^{\infty,\infty}$ are iid $N(0,1)$, characteristic function argument then shows that the limiting process $\sum_{k=1}^{\infty}\xi_{ik}\phi_k(x)$ is Gaussian for $x \in [0,1]$ with mean 0 and covariance $G(x,x')$ defined in (2.2). Our work has gone much beyond the scope of Ma, Yang, and Carroll (2012) which requires that $\kappa < \infty$.

The normality Assumption (A3) on $\{\xi_{ik}\}_{i=1,k=1}^{\infty,\infty}$ cannot be relaxed also because each ξ_{ik} appears in only N_i observations $(X_{ij}, Y_{ij})_{j=1}^{N_i}$, see (2.3). The strong approximation result in Lemma A.3 therefore does not apply since N_i is a random variable with fixed distribution and does not go to infinity. In contrast, Cao, Yang, and Todem (2012) and Cao et al. (2012) do not require the ξ_{ik}'s to be normal for dense functional data.

Next, for a nonnegative integer l and a continuous function $L(x)$, define:

$$\mu_{l,x}(L) = \begin{cases} \int_{-x/h}^{1} v^l L(v)\,dv, & x \in [0,h) \\ \mu_l(L) = \int_{-1}^{1} v^l L(v)\,dv, & x \in [h, 1-h] \\ \int_{-1}^{(1-x)/h} v^l L(v)\,dv, & x \in (1-h, 1] \end{cases} \tag{3.1}$$

$$D_x(L) = \mu_{2,x}(L)\,\mu_{0,x}(L) - \mu_{1,x}^2(L), \tag{3.2}$$

and the equivalent kernel function, see Fan and Gijbels (1996):

$$K_x^*(u) = K(u)\{\mu_{2,x}(K) - \mu_{1,x}(K)u\}D_x^{-1}(K), \quad K_{x,h}^*(u) = K_x^*(u/h)/h \tag{3.3}$$

where $D_x^{-1}(K)$ exists by Lemma A.5. One may verify:

$$\mu_{0,x}(K_x^*) = 1,\ \mu_{1,x}(K_x^*) = 0,\ D_x(K) = \mu_2(K),\quad K_x^*(u) \equiv K(u), \forall x \in [h, 1-h].$$

The asymptotic variance function is

$$\sigma_n^2(x) \overset{\text{def}}{=} \frac{\|K_x^*\|_2^2\sigma_Y^2(x)}{nhf(x)\mathrm{E}N_1}\left[1 + \frac{\mathrm{E}\left(N_1^2 - N_1\right)}{\mathrm{E}N_1}\frac{G(x,x)f(x)h}{\sigma_Y^2(x)\|K_x^*\|_2^2} + \frac{\mu_{1,x}\left(K_x^{*2}\right)\left\{\sigma_Y^2(x)f(x)\right\}'h}{\|K_x^*\|_2^2\sigma_Y^2(x)f(x)}\right]. \tag{3.4}$$

Define $z_{1-\alpha/2} \overset{\text{def}}{=} \Phi^{-1}(1-\alpha/2)$ and

$$Q_h(\alpha) \overset{\text{def}}{=} a_h + a_h^{-1}[\log\{\sqrt{C(K)}/(2\pi)\} - \log\{-\log\sqrt{1-\alpha}\}] \tag{3.5}$$

with $a_h = \sqrt{-2\log h}$, $C(K) = \{\int_{-1}^{1} K'(x)^2 dx\}\{\int_{-1}^{1} K^2(x)dx\}^{-1}$.

Theorem 3.1. Under Assumptions (A1)–(A6), for any $\alpha \in (0,1)$

$$\lim_{n\to\infty} \mathrm{P}\{\sup_{x\in[0,1]}|\widehat{m}(x) - m(x)|/\sigma_n(x) \le Q_h(\alpha)\} = 1-\alpha,$$
$$\lim_{n\to\infty} \mathrm{P}\{|\widehat{m}(x) - m(x)|/\sigma_n(x) \le z_{1-\alpha/2}\} = 1-\alpha, \forall x \in [0,1],$$

with $\sigma_n^2(x)$ and $Q_h(\alpha)$ given in (3.4) and (3.5).

By Theorem 3.1, we construct the SCC and SCI for $m(x)$ as follows:

Corollary 3.1. Assume (A1)–(A6). A $100(1-\alpha)\%$ SCC for $m(x)$ is

$$[\widehat{m}(x) \pm \sigma_n(x) Q_h(\alpha)]. \quad (3.6)$$

A shoal of confidence intervals (SCI) is given by

$$[\widehat{m}(x) \pm \sigma_n(x) z_{1-\alpha/2}]. \quad (3.7)$$

A simple approximation of $\sigma_n^2(x)$ is given by

$$\sigma_{n,\mathrm{IID}}^2(x) = \frac{\|K_x^*\|_2^2 \sigma_Y^2(x)}{nhf(x)\mathrm{E}N_1}. \quad (3.8)$$

Proposition 3.1. Given (A2), (A3), and (A6), then $\sup_{x\in[0,1]} |\sigma_n^{-1}(x)\sigma_{n,\mathrm{IID}}(x) - 1| = \mathcal{O}(h)$.

The above proposition allows one to use $\sigma_{n,\mathrm{IID}}^2(x)$ in place of $\sigma_n^2(x)$ for constructing SCC, which is equivalent to asymptotically neglecting the longitudinal dependence structure as a result of sparsity. This phenomenon was previously noted in Wang, Carroll, and Lin (2005), and Ma, Yang, and Carroll (2012), and entails enormous conceptual and computational advantages: for all practical purposes of constructing and using SCC, one can treat a sparse functional data $\{X_{ij}, Y_{ij}\}$, $1 \le j \le N_i$, $1 \le i \le n$ as if it were iid, and apply existing local linear method without additional accommodation for dependence among repeated measurements.

4. IMPLEMENTATION

Now we outline the construction of the SCC and SCI. Recall the definition of $\widehat{m}(x)$. The practical implementation of (3.6) and (3.7) is via choosing the bandwidth and estimating $\mathrm{E}N_1$, $f(x)$ and $\sigma_Y(x)$, see Wang and Yang (2009), and references therein.

For datasets where the range of the X_{ij}'s is [0, 1], a simple rule-of-thumb bandwidth satisfying Assumption (A5) is taken as $h = N_\mathrm{T}^{-1/5}(\log n)^{-1}$ which differs from mean square optimal bandwidth selection but works quite well both in simulations and real examples. In general, one has to first transform the X_{ij}'s so its range becomes [0, 1] and then carry out local linear smoothing with the above bandwidth. An "optimal bandwidth" has never been established in the SCC context, because of the two goals in SCC construction: coverage of the true curve as close as possible to the nominal confidence level, and narrowness of the band, are not quantifiable in a single measure to optimize, such as the mean integrated squared error (MISE). Recent articles on SCC for time series, such as Wu and Zhao (2007), and Zhao and Wu (2008), have used similar undersmoothing bandwidths as ours.

The quantity $\mathrm{E}N_1$ is estimated by N_T/n and the estimator of the density $f(x)$ is

$$\widehat{f}(x) = N_\mathrm{T}^{-1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} K_h(X_{ij} - x).$$

The local linear estimator $\widehat{\sigma}_Y(x) = \widehat{a}_1$ results from

$$(\widehat{a}_1, \widehat{b}_1) = \arg\min_{a_1,b_1} \sum_{i=1}^{n} \sum_{j=1}^{N_i} \left\{\widehat{\varepsilon}_{ij}^2 - a_1 - b_1(X_{ij} - x)\right\}^2 w_{ij},$$

where $\widehat{\varepsilon}_{ij} = Y_{ij} - \widehat{m}(X_{ij})$, $w_{ij} = N_\mathrm{T}^{-1} K_h(X_{ij} - x)$. The consistency of $\widehat{f}(x)$ and $\widehat{\sigma}_Y(x)$ is proved, for example, in Li and Hsing (2010), Yao, Müller, and Wang (2005a). Slutsky's theorem entails that Theorem 3.1 still holds when $\sigma_n(x)$ is replaced by any consistent estimator $\widehat{\sigma}_n(x)$ satisfying that $\|\widehat{\sigma}_n(x) - \sigma_n(x)\|_\infty = o_p(1/\sqrt{\log n})$ as $n \to \infty$. Therefore, the SCC $\widehat{m}(x) \pm \hat{\sigma}_{n,\mathrm{IID}}(x) Q_h(\alpha)$ and the SCI $\widehat{m}(x) \pm \hat{\sigma}_{n,\mathrm{IID}}(x) z_{1-\alpha/2}$ both have asymptotic confidence level $1-\alpha$. It is worthwhile to point out that $\widehat{m}(x)$, $\hat{\sigma}_{n,\mathrm{IID}}(x)$, and $Q_h(\alpha)$, in general, remain stable if the bandwidths slightly vary.

5. MONTE CARLO STUDIES

In what follows, the finite sample performances of the SCCs in this article and that from Ma, Yang, and Carroll (2012, hereafter denoted by LL and MYC, respectively) are compared in terms of uniform coverage rates, average maximal widths as well as graphical visualization. The data are generated as

$$Y_{ij} = m(X_{ij}) + \sum_{k=1}^{\kappa} \xi_{ik}\phi_k(X_{ij}) + \sigma\varepsilon_{ij}. \quad (5.1)$$

One considers first the case of $\kappa = 2$ and $X \sim \mathrm{U}[0, 1]$, $\xi_k \sim \mathrm{N}(0, 1)$, $\varepsilon_{ij} \sim \mathrm{N}(0, 1)$, $m(x) = \sin\{2\pi(x - 1/2)\}$, $\phi_1(x) = -0.4\cos\{\pi(x - 1/2)\}$, $\phi_2(x) = 0.1\sin\{\pi(x - 1/2)\}$, $\sigma = 0.5$ or 1, while $N_i \sim \mathrm{U}\{2, 3, 4\}$, $n = 100, 200, 400, 800$, and the confidence level $1 - \alpha = 0.95, 0.99$.

For the case $\kappa = \infty$, let $\phi_k(x) = \sqrt{\lambda_k}\psi_k(x)$, where $\psi_1(x) = 1/\sqrt{10}$, $\psi_{2k}(x) = \sin(k\pi x)/\sqrt{5}$, $\psi_{2k+1}(x) = \cos(k\pi x)/\sqrt{5}$, $\lambda_k = (1/5)^{2[k/2]}$, $k = 1, 2, \ldots, \infty$. The infinite series $G(x, x') = \sum_{k=1}^{\infty} \phi_k(x)\phi_k(x')$ is well approximated by finite sum $G(x, x') = \sum_{k=1}^{\kappa} \phi_k(x)\phi_k(x')$ where $\kappa = 2001, 3001, 4001, 5001$, as the fraction of variance explained (FVE) criteria, $\mathrm{FVE}_{2001} = \sum_{k=1}^{2001} \lambda_k / \sum_{k=1}^{\infty} \lambda_k > 1 - 10^{-50}$, see Yao, Müller, and Wang (2005b). Other simulation settings remain the same as for $\kappa = 2$.

Figure 1 plots the data ($n = 100, 200$, $\kappa = 2$), the 95% LL SCCs and MYC SCCs, respectively. Table 1 reports the uniform coverage rates of the SCCs from 500 replications, while the average maximal widths of the SCCs are summarized in Table 2. Clearly, the LL SCC outperforms the MYC SCC in all aspects that matter: it is much narrower with coverage frequencies much closer to nominal levels and highlights more sharply the features of the mean function, thus it presents enhanced visual impression and has greater power for testing hypotheses. These hold true regardless of sample size n, confidence level $1 - \alpha$, noise level σ, and the number κ of positive eigenvalues. Results in Tables 1 and 2 are also evidence that the simple standard deviation $\hat{\sigma}_{n,\mathrm{IID}}(x)$ is a viable substitute of the theoretical standard deviation $\sigma_n(x)$ that accounts for the (asymptotically negligible) dependence among repeated measurements, as all SCCs have been computed using $\hat{\sigma}_{n,\mathrm{IID}}(x)$.

One referee has pointed out that confidence bands based on extreme value theory typically have poor finite sample properties (in contrast to bootstrap methods), for which there is also

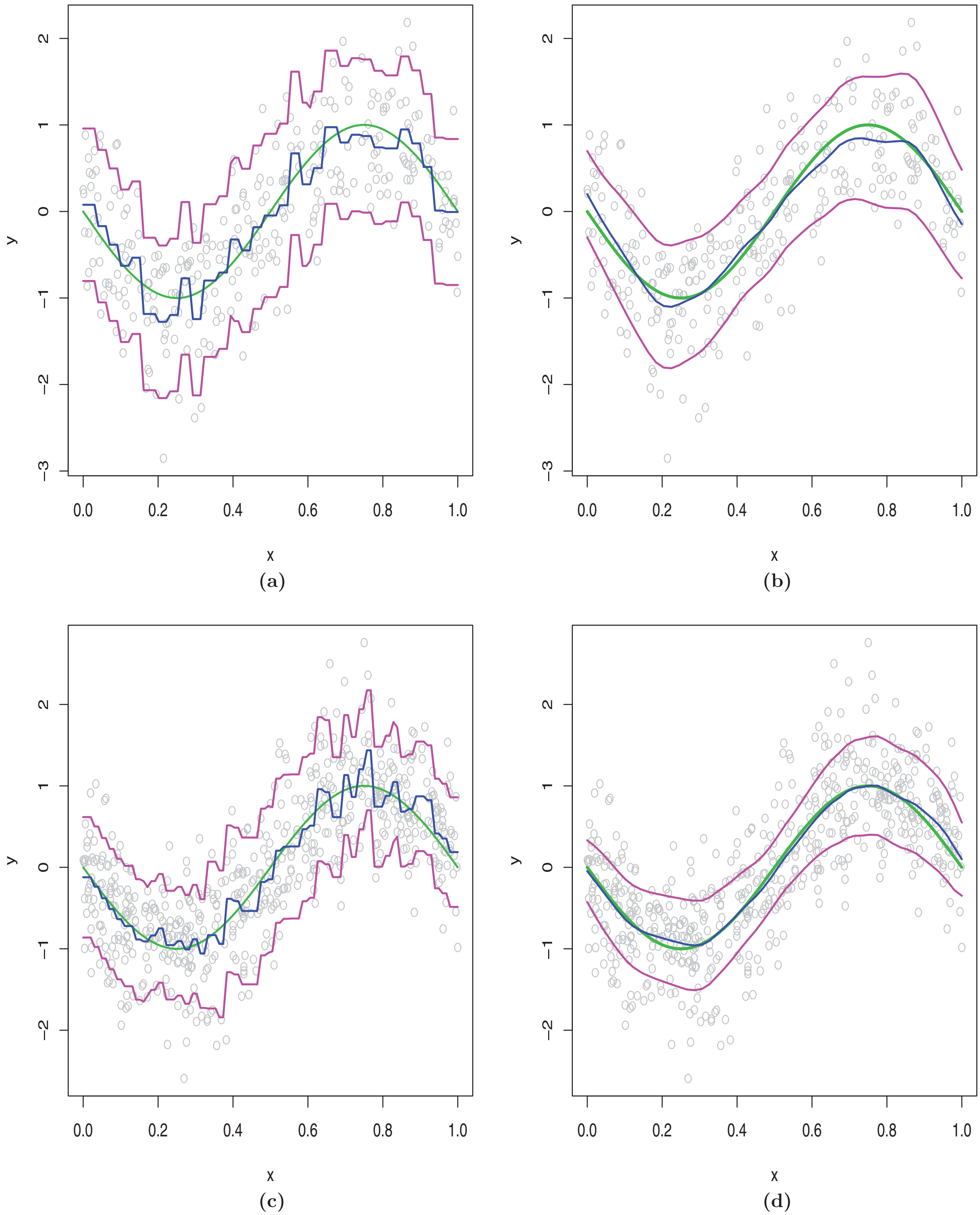

Figure 1. Plots of the 95% SCCs (upper and lower lines), data (circle), true mean (thick median line), and mean estimator (thin median line) when $\sigma = 0.5$: (a) $n = 100$, MYC; (b) $n = 100$, LL; (c) $n = 200$, MYC; (d) $n = 200$, LL.

Table 1. Uniform coverage rates of SCCs from 500 replications: "∗∗"/"∗" if LL is better than/as good as MYC

	$\sigma = 0.5$								$\sigma = 1.0$							
$\kappa = 2$	$n = 100$		$n = 200$		$n = 400$		$n = 800$		$n = 100$		$n = 200$		$n = 400$		$n = 800$	
$1-\alpha$	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990
MYC	0.912	0.938	0.930	0.966	0.942	0.978	0.952	0.998	0.884	0.922	0.920	0.960	0.938	0.980	0.962	0.996
LL	0.946	0.974	0.954	0.974	0.956	0.988	0.952	0.992	0.906	0.966	0.924	0.966	0.944	0.984	0.954	0.996
Comparison	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗
"$\kappa = \infty$"	$\kappa \approx 2001$		$\kappa \approx 3001$		$\kappa \approx 4001$		$\kappa \approx 5001$		$\kappa \approx 2001$		$\kappa \approx 3001$		$\kappa \approx 4001$		$\kappa \approx 5001$	
$1-\alpha$	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990
MYC	0.896	0.958	0.926	0.968	0.964	0.980	0.944	0.986	0.892	0.932	0.922	0.950	0.938	0.980	0.958	0.996
LL	0.926	0.996	0.968	0.984	0.962	0.992	0.948	0.994	0.898	0.942	0.964	0.972	0.956	0.982	0.954	0.992
Comparison	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗	∗∗

Table 2. Average maximal widths of the SCCs from 500 replications: "∗" if LL is narrower than MYC

	$\sigma = 0.5, \kappa = 2$				$\sigma = 1, \kappa = 2$				$\sigma = 0.5, \kappa = \infty$				$\sigma = 1, \kappa = \infty$			
n	100	200	400	800	100	200	400	800	100	200	400	800	100	200	400	800
								$1-\alpha = 0.95$								
MYC	2.69	2.09	1.62	1.30	3.38	2.64	2.22	1.82	2.24	1.88	1.57	1.12	2.82	2.33	2.04	1.64
LL	2.47	1.95	1.56	1.13	3.07	2.43	1.91	1.57	2.02	1.62	1.30	1.01	2.70	2.11	1.78	1.49
Comparison	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗
								$1-\alpha = 0.99$								
MYC	3.10	2.41	2.02	1.59	3.91	3.03	2.41	1.98	2.35	2.07	1.69	1.21	3.38	2.68	2.15	1.70
LL	2.99	2.33	1.86	1.51	3.72	2.91	2.29	1.80	2.14	1.94	1.47	1.18	3.27	2.52	1.89	1.63
Comparison	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗	∗

theoretical support, see Hall (1993). The main reason finite sample behavior is so much better in our setting (Tables 1 and 2) is the large total sample size, $N_{\mathrm{T}} \approx 3n = 300, 600, 1200, 2400$. For smaller sample size $n = 20, 50, 80$, the SCCs become much wider and have much poorer coverage rates, see Table 3. We note, however, that the growth curve data briefly discussed in Section 1 have sample size $n = 132$ and total sample size $N_{\mathrm{T}} = 310$ (all the N_i's range between 2 and 4), which closely resembles our examples with $n \geq 100$. Another example, the CD4 cell counts data in Yao, Müller, and Wang (2005b), has $n = 283$ and N_i's range between 1 and 14, with a median of 6. Therefore, the superior performance of our proposed SCC in Tables 1 and 2 is to be expected in most commonly encountered sparse functional data.

Following one referee's suggestion, a MISE-relevant undersmoothing bandwidth fulfilling Assumption (A5) $h = h_{\mathrm{opt}}(\log n)^{-0.25}$ is also examined, where h_{opt} is the MISE optimal bandwidth with order $N_T^{-1/5}$, see Fan and Gijbels (1996). Table 4 reports the uniform coverage rates and the average maximal widths of the SCCs from 500 replications, showing coverage rates comparable with those obtained in Table 1, but increased width relative to Table 2.

Table 3. SCCs for sample size $n = 20, 50, 80$ from 500 replications

	$\sigma = 0.5$						$\sigma = 1.0$					
	$n = 20$		$n = 50$		$n = 80$		$n = 20$		$n = 50$		$n = 80$	
						$\kappa = 2$						
$1-\alpha$	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990
Coverage	0.838	0.904	0.882	0.922	0.908	0.942	0.788	0.878	0.872	0.926	0.882	0.938
Width	4.02	5.07	3.09	3.80	2.61	3.16	5.05	6.38	3.81	4.69	3.24	3.94
						$\kappa = \infty$						
$1-\alpha$	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990
Coverage	0.862	0.902	0.894	0.932	0.904	0.960	0.812	0.902	0.824	0.912	0.876	0.936
Width	3.33	4.21	2.54	3.63	2.48	3.09	4.46	5.64	3.46	4.25	3.07	3.51

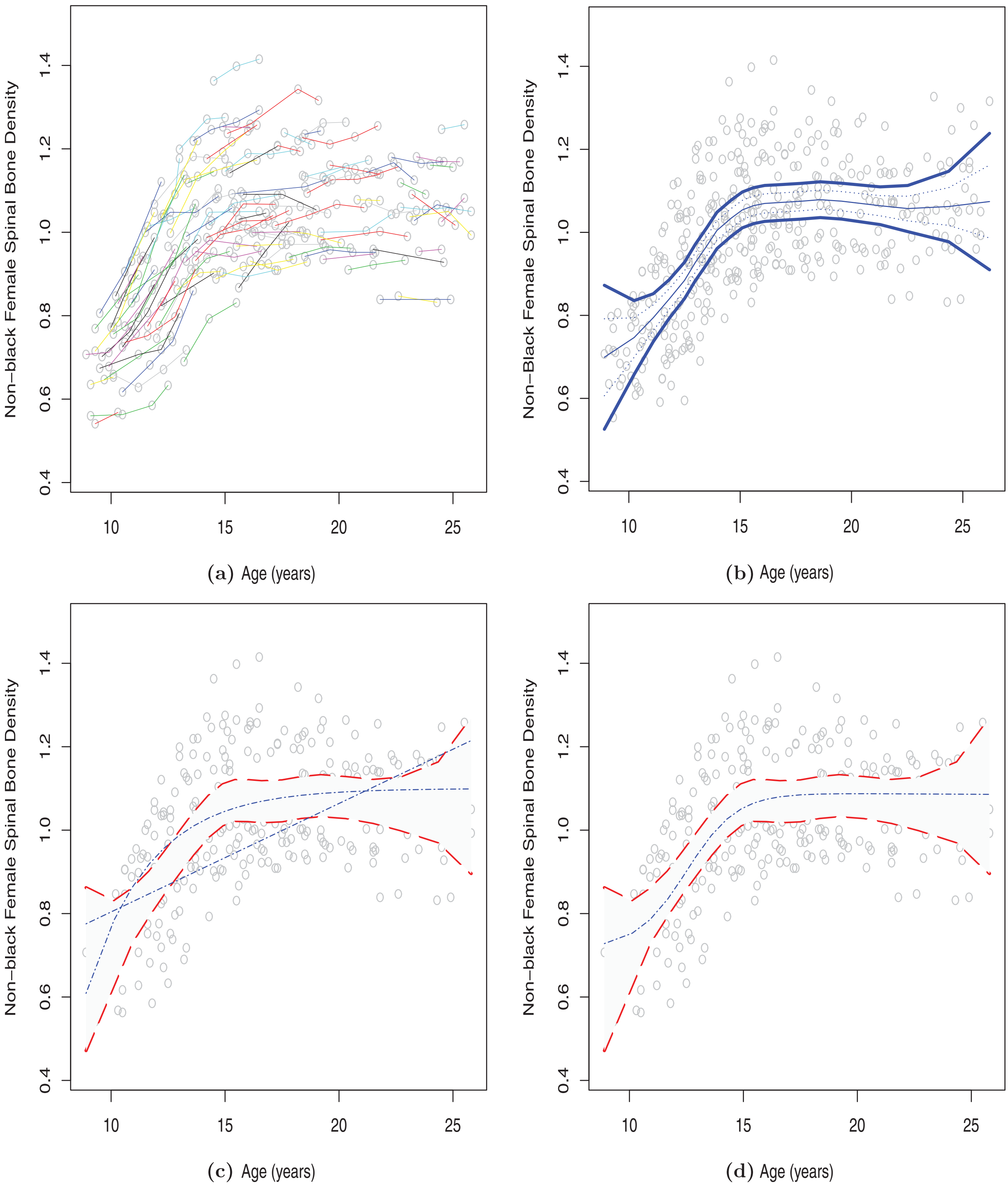

Figure 2. Plots of the nonblack female growth curve: (a) the data; (b) local linear estimator (median solid line), 90% SCC (thick lines) and SCI (dotted lines); (c) 90% SCC (dashed lines), linear growth model (straight line) and monomolecular growth model (dotted curve); (d) 90% SCC (dashed lines) and logistic growth model (dotted curve).

Table 4. SCCs with bandwidth $h = h_{\text{opt}} (\log n)^{-0.25}$ from 500 replications

	$\sigma = 0.5$								$\sigma = 1.0$							
	$n = 100$		$n = 200$		$n = 400$		$n = 800$		$n = 100$		$n = 200$		$n = 400$		$n = 800$	
								$\kappa = 2$								
$1-\alpha$	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990
Coverage	0.958	0.978	0.956	0.980	0.954	0.990	0.956	0.994	0.942	0.964	0.948	0.970	0.954	0.986	0.954	0.998
Width	3.52	4.16	2.75	3.15	2.17	2.56	1.55	1.93	4.17	4.95	3.24	3.73	2.57	2.78	1.79	2.10
								$\kappa = \infty$								
$1-\alpha$	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990	0.950	0.990
Coverage	0.932	0.998	0.968	0.986	0.966	0.994	0.952	0.998	0.936	0.944	0.964	0.976	0.956	0.986	0.952	0.992
Width	3.08	4.07	2.40	2.63	1.95	2.11	1.42	1.56	3.71	4.43	2.93	3.30	1.41	2.52	1.77	2.05

The above studies have convincingly demonstrated the superior numerical properties of the LL SCC, which will be used to analyze real data in the next section.

6. APPLICATION

We now return to the growth curve data introduced in Section 1, consisting of measurements of spinal bone mineral density at various ages for $n = 132$ nonblack females (total size $N_T = 310$). They are a subset of the data studied in Bachrach et al. (1999) and James, Hastie, and Sugar (2000), and it was claimed in Bachrach et al. (1999) that the individuals were independently surveyed from different biophysical and social environments.

As mentioned before, our estimate of the mean spinal bone density curve increases with age, with highest growth rate during early adolescence (9–15 yr) and peaks during late puberty (16–26 yr), see, Figure 2(b). Although these visual impressions seem to validate what is known in pediatrics, one referee cautions that they are best tested with the aid of SCC. We have carried out such testing as follows.

To test the hypotheses:

$$H_0 : m(x) = m_0(x) \quad \text{versus}$$
$$H_a : m(x) \neq m_0(x) \quad \text{for some } x \in [0, 1] \tag{6.2}$$

for a given function $m_0(x)$, one applies the rule "Reject H_0 if the curve $\{(x, m_0(x)), x \in [0, 1]\}$ is not covered entirely by the asymptotic $100(1-\alpha)\%$ SCC", with asymptotic power α under H_0, 1 under H_a due to Theorem 3.1. Three candidate monotone growth functions are tested: the "linear growth function" $m_{0,1}(x)$, the "monomolecular growth function" $m_{0,2}(x)$ (a: growth size, b : shape and k: rate) and the "logistic growth function" $m_{0,3}(x)$ (a_1: growth size, a_2: shape, a_3: rate and, similarly, b_1: growth size, b_2: rate and b_3: puberty effects):

$$\begin{aligned} m_{0,1}(x) &= c + \beta x, \\ m_{0,2}(x) &= a\{1 - b\exp(-kx)\}, \\ m_{0,3}(x) &= a_1/\{1 + a_2 \exp(-a_3 x)\} \\ &\quad + b_1/\{1 + \exp(-b_2(x - b_3))\}, \end{aligned} \tag{6.3}$$

see Sluis et al. (2002), and references therein. Figure 2(c) overlays the 90% SCC and the functions $m_{0,1}(x)$ and $m_{0,2}(x)$ obtained via nonlinear least squares regression, while Figure 2(d) does the same for $m_{0,3}(x)$. Clearly, at the 10% significance level, both $m_{0,1}(x)$ and $m_{0,2}(x)$ are rejected, while the logistic growth model represented by $m_{0,3}(x)$ is accepted. Our finding reconfirms the selected model in Sluis et al. (2002).

An R algorithm scbsfda.R of our method has been provided on *www.quantlet.org.*

7. DISCUSSION

In this article, a local linear SCC is proposed for the mean function of sparse functional data. In terms of theoretical properties such as smoothness and narrowness, as well as computation ease and conceptual appeal, it is comparable to existing methods for iid data, such as Johnston (1982), Fan and Zhang (2000), and Claeskens and Van Keilegom (2003). The main advantage of our method is to have extended these desirable properties from iid data to sparse functional data, by relying heavily on complicated Kac-Rice Formula and Double Sum Method to show the absolute maximum of a sequence of nonstationary Gaussian processes (the maximal deviation of the local linear estimator for sparse functional mean) converges to the standard Gumbel distribution. These new additions to the nonparametric smoothing toolkit will definitely be used in other settings where "Hungarian embedding" fails.

Our extensive Monte Carlo experiments show that the convergence of maximal deviation to the Gumbel distribution is rather rapid for sparse functional data with total sample size 300 or larger, which is typical in application. Both theoretical and simulation results have also demonstrated convincingly that dependence among repeated measurements has little effect, and one can practically compute the local linear SCC as if the data are iid. With these observations in mind, we have applied the SCC methodology to test several growth models for a growth curve data.

APPENDIX A

A.1 Preliminaries

We introduce Lemmas A.1–A.4 for the proof of Theorem 3.1 (Appendix A.2). For the details of Lemma A.1, see Cierco-Ayrolles, Croquette, and Delmas (2003) or the online supplementary materials.

Lemma A.1. [Cierco-Ayrolles, Croquette, and Delmas (2003)] Let $X(t)$ be a Gaussian process with almost surely $\mathcal{C}^1$ sample paths on

$[0, T]$ and, for any integer a, $a^{[2]} = a(a-1)$. Then

$$\begin{aligned}&\mathrm{P}\{|X(0)| > u\} + \mathrm{E}\left[\left(U_u^X[0,T] + D_{-u}^X[0,T]\right) I_{\{|X(0)|\leqslant u\}}\right] \\ &\quad - \frac{1}{2}\mathrm{E}\left(U_u^X[0,T] + D_{-u}^X[0,T]\right)^{[2]} \\ &\quad \le \mathrm{P}\left\{\sup_{x\in[0,T]}|X(t)| > u\right\} \le \mathrm{P}\{|X(0)| > u\} \\ &\quad + \mathrm{E}\left[\left(U_u^X[0,T] + D_{-u}^X[0,T]\right) I_{\{|X(0)|\leqslant u\}}\right]. \qquad \text{(A.1)}\end{aligned}$$

Lemma A.2. [Theorem 1 of Cierco-Ayrolles, Croquette, and Delmas (2003)] Suppose X is a $\mathcal{C}^1$ real-valued Gaussian process defined on an interval I and $\{X(t), X(s), X'(t), X'(s)\}$ is nondegenerate $\forall t \neq s, (t,s) \in I^2$. Then, denoting p_V the probability density of a random vector V:

$$\begin{aligned}&\mathrm{E}\left(U_u^X[I]^{[2]}\right) \\ &= \int_{I^2}\int_{(0,\infty)^2} |x_1'||x_2'| p_{X_t;X_s;X_t';X_s'}(u;u;x_1';x_2')dx_1'dx_2'dtds, \\ &\mathrm{E}\left(U_u^X[I]\, D_{-u}^X[I]\right) \\ &= \int_{I^2}\int_0^{+\infty}\int_{-\infty}^0 |x_1'||x_2'| p_{X_t;X_s;X_t';X_s'}(u;-u;x_1';x_2')dx_1'dx_2'dtds.\end{aligned}$$

Lemma A.3. [Theorem 2.6.7 of Csőrgő and Révész (1981)] Suppose that ξ_i, $1 \le i \le n$ are iid with $\mathrm{E}\xi_1 = 0$, $\mathrm{E}\xi_1^2 = 1$, and $H(x) > 0$ $(x \ge 0)$ is an increasing continuous function such that $x^{-2-\gamma}H(x)$ is increasing for some $\gamma > 0$ and $x^{-1}\log H(x)$ is decreasing with $\mathrm{E}H(|\xi_1|) < \infty$. Then, there exist constants $C_1, C_2, a > 0$ which depend only on the distribution of ξ_1 and a sequence of Brownian motions $\{W_n(t), 0 \le t < \infty\}_{n=1}^\infty$ such that for any $\{x_n\}_{n=1}^\infty$ satisfying $H^{-1}(n) < x_n < C_1(n\log n)^{1/2}$ and $S_k = \sum_{i=1}^k \xi_i$

$$\mathrm{P}\{\max_{1\le k\le n}|S_k - W_n(k)| > x_n\} \le C_2 n\{H(ax_n)\}^{-1}.$$

Lemma A.4. [Theorem 1.2 of Bosq (1996)] Suppose that ξ_i, $1 \le i \le n$ are iid with $\sigma^2 = \mathrm{E}\xi_1^2$, $\mathrm{E}\xi_1 = 0$ and there exists $c > 0$ such that for $r = 3, 4, \ldots$, $\mathrm{E}|\xi_1|^r \le c^{r-2} r!\,\mathrm{E}\xi_1^2 < +\infty$, then for each $n > 1, t > 0$, $\mathrm{P}(|S_n| \ge \sqrt{n}\sigma t) \le 2\exp\{-t^2(4 + 2ct/\sqrt{n}\sigma)^{-1}\}$.

A.2 Proof of Theorem 3.1

We begin with an outline of four main ingredients in the proof of Theorem 3.1.

(I) Denote by $\mathbf{m} = (m(X_{ij}))$, $\varepsilon = (\sigma(X_{ij})\varepsilon_{ij})$, $\xi = (\sum_{k=1}^\infty \xi_{ik}\phi_k(X_{ij}))$ the signal, noise and principal component vectors in the decomposition $\mathbf{Y} = \mathbf{m} + \xi + \varepsilon$, then

$$\widehat{m}(x) - m(x) = \widetilde{m}(x) - m(x) + \widetilde{e}(x), \widetilde{e}(x) = \widetilde{\xi}(x) + \widetilde{\varepsilon}(x), \qquad \text{(A.2)}$$

where $\widetilde{\xi}(x) = e_0^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{W}\xi$ and $\widetilde{\varepsilon}(x) = e_0^\mathsf{T}(\mathbf{X}^\mathsf{T}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{W}\varepsilon$. The error structure in (A.2) allows one to study the asymptotics of $\sup_{x\in[0,1]}|\{\widetilde{m}(x) - m(x)\}/\sigma_n(x)|$ (the bias) and $\sup_{x\in[0,1]}|\widetilde{e}(x)/\sigma_n(x)|$ (the error) separately in Lemmas A.6-A.14, with $\sigma_n(x)$ given in (3.4).

(II) The strong approximation of Lemma A.9 shows that the error term $\widetilde{e}(x)/\sigma_n(x)$ is equivalent to a Gaussian process $\eta_n(x)$ defined in (A.9).

(III) Two advanced probability tools: the Kac–Rice Formula (Lemmas A.1–A.2) and the Double Sum Method of Piterbarg (1996) are combined to establish Lemmas A.11–A.13 . Partitioning a growing interval into small and large subintervals, the Double Sum Method whitens the process $\eta_n(x)$ while the Kac–Rice Formula allows investigating the local asymptotics of $\eta_n(x)$ on each subinterval.

(IV) The limiting distribution of $\sup_{x\in[0,1]}|\eta_n(x)|$, negligibility of the bias, and the Slutsky's Theorem prove Theorem 3.1.

We next introduce some notations used throughout. For functions $a_n(x)$ and $b_n(x)$, $a_n(x) = \scriptstyle{\mathcal{U}}\{b_n(x)\}$ and $a_n(x) = \mathcal{U}\{b_n(x)\}$ respectively means that, as $n \to \infty$, $\sup_{x\in[0,1]}|a_n(x)/b_n(x)| = o(1)$ and $\sup_{x\in[0,1]}|a_n(x)/b_n(x)| = \mathcal{O}(1)$. In addition, $a_n(x) = \scriptstyle{\mathcal{U}}_{\mathrm{a.s.}}\{b_n(x)\}$ and $a_n(x) = \mathcal{U}_{\mathrm{a.s.}}\{b_n(x)\}$ respectively means that, as $n \to \infty$, $a_n(x) = \scriptstyle{\mathcal{U}}\{b_n(x)\}$ and $a_n(x) = \mathcal{U}\{b_n(x)\}$ almost surely, and $o_{\mathrm{a.s.}}$, o_p, $\mathcal{O}_{\mathrm{a.s.}}$, $\mathcal{O}_p$ are similarly defined. Further denote

$$\mathbf{D}_x = \begin{pmatrix} \mu_{2,x}(K) & -\mu_{1,x}(K) \\ -\mu_{1,x}(K) & \mu_{0,x}(K)\end{pmatrix}, \qquad \text{(A.3)}$$

with $\mu_{l,x}(K)$ given in (3.1)

$$\widehat{\varepsilon}(x) = f^{-1}(x) N_\mathrm{T}^{-1}\sum_{i,j} K_{x,h}^*(X_{ij} - x)\sigma(X_{ij})\varepsilon_{ij}, \qquad \text{(A.4)}$$

$$\widehat{\xi}(x) = f^{-1}(x) N_\mathrm{T}^{-1}\sum_{i,j} K_{x,h}^*(X_{ij} - x)\sum_{k=1}^\infty \phi_k(X_{ij})\xi_{ik},$$

with $K_{x,h}^*(u)$ given in (3.3)

$$R_{ij,\varepsilon}(x) = K_{x,h}^*(X_{ij} - x) D_x(K)\sigma(X_{ij}), \qquad \text{(A.5)}$$

$$R_{i,\xi} = \sum_{j=1}^{N_i} K_{x,h}^*(X_{ij} - x) D_x(K)\sum_{k=1}^\infty \phi_k(X_{ij}),$$

with $D_x(K)$ given in (3.2)

$$\sigma_{\varepsilon,n}^2(x) = f^{-2}(x) N_\mathrm{T}^{-2} D_x^{-2}(K)\sum_{i,j} R_{ij,\varepsilon}^2(x), \qquad \text{(A.6)}$$

$$\sigma_{\xi,n}^2(x) = f^{-2}(x) N_\mathrm{T}^{-2} D_x^{-2}(K)\sum_{i=1}^n R_{i,\xi}^2(x),$$

with $K_x^{*\prime}(x) = dK_x^*(x)/dx$, $\mu_{l,x}(L)$ given in (3.1)

$$C_x(K) = \frac{\mu_{0,x}\{K_x^{*\prime}(x)^2\}}{\mu_{0,x}\{K_x^*(x)^2\}} - \frac{\mu_{0,x}^2\{K_x^*(x) K_x^{*\prime}(x)\}}{\mu_{0,x}^2\{K_x^*(x)^2\}}. \qquad \text{(A.7)}$$

It is easily verified that $C_x(K) = C(K)$, $\forall x \in [h, 1-h]$ with $C(K)$ given in (3.5).

Lemma A.5. Under Assumptions (A5)–(A6), for $x \in [0,1]$

$$0 < D_0(K) \le D_x(K) \le D_{1/2}(K) = \mu_2(K) < +\infty, \qquad \text{(A.8)}$$

while $\sup_{x\in[0,1]}|C_x(K)| < \infty$.

Proof. See the online supplementary materials. □

Lemma A.6. Under Assumptions (A1)-(A6), for $D_x(K)$ given in (3.2) and $\mathbf{D}_x$ in (A.3),

$$(\mathbf{X}^\mathsf{T}\mathbf{W}\mathbf{X})^{-1} = f^{-1}(x)\,\mathrm{diag}(1, h^{-1})\left\{D_x^{-1}(K)\,\mathbf{D}_x + \Delta_{1,n}(x)\right\} \times \mathrm{diag}(1, h^{-1})$$

as $n \to \infty$, where the 2×2 random matrices $\Delta_{1,n}(x) = \mathcal{U}(h) + \mathcal{U}_{\mathrm{a.s.}}\{\sqrt{\log n/(nh)}\}$.

Proof. See the online supplementary materials. □

Lemma A.7. Under Assumptions (A1)–(A6), as $n \to \infty$, $\|\widetilde{m}(x) - m(x)\|_\infty = \mathcal{O}_{\mathrm{a.s.}}(h^2)$.

Proof. It simply follows from Lemma A.6 and (A.2). See Proof of Theorem 6.5, p. 268 of Fan and Yao (2005). □

Lemma A.8. Under Assumptions (A1)–(A6), for $\widehat{\varepsilon}(x)$ and $\widehat{\xi}_k(x)$ given in (A.4),

$$\widetilde{e}(x) = \{1 + \Delta_{2,n}(x)\}\{\widehat{\varepsilon}(x) + \widehat{\xi}(x)\}$$

as $n \to \infty$, where the 2×2 random matrices $\Delta_{2,n}(x) = \mathcal{U}(h) + \mathcal{U}_{\mathrm{a.s.}}\{\sqrt{\log n/(nh)}\}$.

Proof. See the online supplementary materials. □

Let $X_{ij}, 1 \le i \le n,\ 1 \le j \le N_i$ be descendingly ordered as $X_{(t)}$, $1 \le t \le N_{\mathrm{T}}$, $S_q = \sum_{t=1}^{q} \varepsilon_{(t)}$ where $\varepsilon_{(t)}$ is corresponding in index to $X_{(t)}$.

Lemma A.9. Given (A1)–(A6), then there exists a sequence of Wiener processes $\{W_{N_{\mathrm{T}}}(t)\}_{t=1}^{N_{\mathrm{T}}}$ independent of $\{N_i,\ X_{ij},\ \xi_i\ 1 \le i \le n, 1 \le j \le N_i, 1 \le k \le \infty\}$ such that as $n \to \infty$ and for some $t^{'} > 2/5$

$$\|\widehat{\varepsilon}(x) - \widehat{\varepsilon}_{N_{\mathrm{T}}}(x)\|_{\infty} = \mathcal{O}_{\mathrm{a.s.}}(n^{-t^{'}}),$$

with $\widehat{\varepsilon}_{N_{\mathrm{T}}}(x) = \{N_{\mathrm{T}} f(x)\}^{-1} \sum_{t=1}^{N_{\mathrm{T}}} K^{*}_{x,h}(X_{(t)} - x)\sigma(X_{(t)})\{W_{N_{\mathrm{T}}}(t) - W_{N_{\mathrm{T}}}(t-1)\}$.

Proof. See the online supplementary materials. □

Lemma A.10. Under Assumptions (A1)–(A6), as $n \to \infty$,

$$\left\| N_{\mathrm{T}}^{-1} \sum_{i,j} R^2_{ij,\varepsilon}(x) - \mathrm{E}\, R^2_{1,\varepsilon}(x) \right\|_{\infty} = \mathcal{O}_{\mathrm{a.s.}}\{\sqrt{\log n/(nh)}\},$$

$$\left\| N_{\mathrm{T}}^{-1} \sum_{i=1}^{n} R^2_{i,\xi}(x) - (\mathrm{E}\, N_1)^{-1}\, \mathrm{E}\, R^2_{1,\xi}(x) \right\|_{\infty} = \mathcal{O}_{\mathrm{a.s.}}\{\sqrt{\log n/(nh)}\},$$

with $R_{ij,\varepsilon}(x)$ and $R_{ik,\xi_k}(x)$ given in (A.5).

Proof. See the online supplementary materials. □

Throughout the remainder, define the standardized noise processes as

$$\eta_n(x) = \eta(x) = \{\widehat{\varepsilon}_{N_{\mathrm{T}}}(x) + \widehat{\xi}(x)\}\left\{\sigma^2_{\varepsilon,n}(x) + \sigma^2_{\xi,n}(x)\right\}^{-1/2}, x \in [0,1] \tag{A.9}$$

with $\widehat{\varepsilon}_{N_{\mathrm{T}}}(x)$, $\widehat{\xi}(x)$, $\sigma^2_{\varepsilon,n}(x)$ and $\sigma^2_{\xi,n}(x)$, respectively, given in Lemma A.9, (A.4) and (A.6). For any n and fixed x ,

$$\mathcal{L}\left\{\eta(x)\,|(X_{ij}, N_i), 1 \le j \le N_i, 1 \le i \le n\right\} = \mathrm{N}(0,1),$$

and hence $\mathcal{L}\{\eta(x)\} = \mathrm{N}(0,1)$ which implies $\eta(x)$ is a standardized Gaussian process (nonstationary).

To compute the extreme value of $\eta(x)$, one needs to study its correlation function. In the following, denote $xh^{-1} = t \in [0, h^{-1}]$, $m_t = m(t) = \mathrm{E}\,\eta(t)$, $r(t,s) = \mathrm{E}\,\eta(t)\eta(s)$, $r_t = r(t,t)$, $r_{0t} = r(0,t)$, $r_{1,0}(t,s) = \partial r(\alpha,\beta)/\partial\alpha|_{(t,s)}$, $r_{1,1}(t,s) = \partial^2 r(\alpha,\beta)/\partial\alpha\partial\beta|_{(t,s)}$, $\eta_{1,1}(t,s) = \partial\, \mathrm{E}\,\eta(t)\eta(s)/\partial t \partial s$, $t, s \in [0, h^{-1}]$ and $C(t) \overset{\mathrm{def}}{=} C_{th}(K), t \in [0, h^{-1}]$, with $C_{th}(K)$ as in (A.7), so that $C(t) \equiv C(K)\,, \forall t \in [1, h^{-1} - 1]$. Clearly, for any n,

$$m(t) = 0, r(t,t) = r_t \equiv 1. \tag{A.10}$$

and it is easy to verify that for $\forall t \in [0, h^{-1}]$

$$r_{1,0}(t,t) = 0, \tag{A.11}$$

while for $v^2 = \mathrm{var}\{\eta'(t)|\eta(0), \eta(t)\}$, see (15) in the online supplementary materials, $s, t \in [0, h^{-1}]$ and $|t - s| \ge 2$,

$$r_{st} = r_{1,0}(t,s) = 0,\ v^2 = r_{1,1}(t,t). \tag{A.12}$$

Lemma A.11. Under Assumptions (A1)–(A6)

$$\lim_{n\to\infty} \sup_{t\in[0,h^{-1}]} |r_{1,1}(t,t) - C(t)| = 0. \tag{A.13}$$

There exist constants $0 < c < C < \infty$, $1 > \delta > 0$, such that for large n

$$\inf_{t,s\in[0,h^{-1}], |t-s|<2} r(t,s) \ge -1 + c > -1,$$
$$\sup_{2>|t-s|\ge\delta,\ t,s\in[0,h^{-1}]} r(t,s) \le 1 - c < 1, \tag{A.14}$$

$$\sup_{0<|t-s|<\delta,t,s\in[0,h^{-1}]} \max[r_{1,0}(t,s)/(t-s),\ \{1 - r^2(t,s)\}/(t-s)^2] \le C, \tag{A.15}$$
$$\inf_{0<|t-s|<\delta,t,s\in[0,h^{-1}]} \min[r_{1,0}(t,s)/(t-s),\ \{1 - r^2(t,s)\}/(t-s)^2] \ge c,$$

$$\sup_{0<|t-s|<\delta,t,s\in[0,h^{-1}]} \frac{r_{1,1}(t,t) - r^2_{1,0}(t,s)/(1-r^2)}{(t-s)^2} \le C, \tag{A.16}$$
$$\inf_{0<|t-s|<\delta,t,s\in[0,h^{-1}]} \frac{r_{1,1}(t,t) - r^2_{1,0}(t,s)/(1-r^2)}{(t-s)^2} \ge c,$$

$$\sup_{|t-s|<2,t,s\in[0,h^{-1}]} |r^2_{1,0}(t,s)/\{1 - r^2(t,s)\}| \le C \tag{A.17}$$
$$\inf_{|t-s|<2,t,s\in[0,h^{-1}]} \frac{|r_{1,0}(t,s)/(1+r)|}{\sqrt{r_{1,1}(t,t) - r^2_{1,0}(t,s)/(1-r^2)}} \ge c$$

Proof. See the online supplementary materials. □

In what follows, the "double sum" method of Piterbarg (1996) will be applied to study the extreme value distribution of the sequence of Gaussian processes $\eta(t)$ over the growing interval $[0, h^{-1}]$. Partition the interval $[1, h^{-1} - 1]$ as $1 = a_1 < b_1 < a_2 < b_2 < \cdots < a_N < b_N = h^1 - 1$, assuming $I_l = [a_l, b_l], l = 1, \ldots, N$, $I'_l = [b_l, a_{l+1}], l = 1, \ldots, N-1$ and the length of I_l and I'_l are λ_n and 2, respectively, where $(\lambda_n + 2)N = h^{-1}$ and $\lambda_n \to \infty$, $N \to \infty$ as $n \to \infty$.

Lemma A.12. Under Assumptions (A1)–(A6), for $u = u_n$ satisfying $2\sqrt{C(K)}N\lambda_n\varphi(u_n)\varphi(0) \to -\log(1-\alpha)$ with $C(K)$ given in (3.5)

$$\lim_{n\to\infty} \mathrm{P}\{\sup\nolimits_{t\in[0,1)\cup_{l=1}^{N-1} I'_l \cup(h^{-1}-1,h^{-1}]} |\eta(t)| \le u\} = 1.$$

Proof. In Lemma A.1, for $\forall\ [a,b] \subseteq [0, h^{-1}]$, one computes according to Cierco-Ayrolles, Croquette, and Delmas (2003) or the online supplementary materials.

$$\begin{aligned}
&\mathrm{E}\left[\left(U^{\eta}_{u}[a,b] + D^{\eta}_{-u}[a,b]\right) I_{\{|X(a)|\le u\}}\right] \\
&= 2\varphi(u)\left\{\varphi(0)\int_a^b \sqrt{r_{1,1}(t,t)}dt\right. \\
&\quad - \int_a^b \left(\varphi(0)\sqrt{r_{1,1}(t,t)}\left[1 - \Phi\left\{\sqrt{r_{1,1}(t,t)}\sqrt{\frac{1-r_{at}}{1+r_{at}}}\frac{u}{v}\right\}\right]\right. \\
&\quad \left. + \frac{r_{1,0}(t,a)}{\sqrt{1-r^2_{at}}}\varphi\left(\sqrt{\frac{1-r_{at}}{1+r_{at}}}u\right)\Phi\left\{\frac{r_{1,0}(t,a)}{1+r_{at}}\frac{u}{v}\right\}\right)dt \\
&\quad - \int_a^b \left(\varphi(0)\sqrt{r_{1,1}(t,t)}\left[1 - \Phi\left\{\sqrt{r_{1,1}(t,t)}\sqrt{\frac{1+r_{at}}{1-r_{at}}}\frac{u}{v}\right\}\right]\right. \\
&\quad \left.\left. - \frac{r_{1,0}(t,a)}{\sqrt{1-r^2_{at}}}\varphi\left(\sqrt{\frac{1+r_{at}}{1-r_{at}}}u\right)\Phi\left\{\frac{-r_{1,0}(t,a)}{1-r_{at}}\frac{u}{v}\right\}\right)dt\right\}.
\end{aligned} \tag{A.18}$$

According to (A.13) and (A.17), it is clear that as $n \to \infty$,

$$\sup_{1\le l\le N-1} \mathrm{E}\left[\left(U^{\eta}_{u}[b_l, b_l+2] + D^{\eta}_{-u}[b_l, b_l+2]\right) I_{\{|X(b_l)|\le u\}}\right] = \mathcal{O}\{\varphi(u)\}.$$

Hence, the upperbound of (A.1) shows that, if $2\sqrt{C(K)}\,N\lambda_n\varphi(u_n)\varphi(0) \to -\log(1-\alpha)$ as $n \to \infty$,

$$\sum_{l=1}^{N-1} \mathrm{P}\left\{\sup\nolimits_{I'_{l'}} |\eta(t)| > u\right\} = \mathcal{O}\left[2N\{1 - \Phi(u)\}\right] + \mathcal{O}\{N\varphi(u)\} = \mathcal{o}(1),$$

Similarly, while $t \in [0,1) \cup (h^{-1} - 1, h^{-1}]$, one can show that

$$\mathrm{P}\{\sup\nolimits_{t\in[0,1)\cup(h^{-1}-1,h^{-1}]} |\eta(t)| > u\} = \mathcal{O}\{1 - \Phi(u)\} + \mathcal{O}\{\varphi(u)\} = \mathcal{o}(1).$$

Finally, this lemma is proved by

$$\mathrm{P}\{\sup_{t\in[0,1)\cup_{l'=1}^{N} I_{l'}\cup(h^{-1}-1,h^{-1}]}|\eta(t)|>u\}$$
$$\le \mathrm{P}\{\sup_{t\in[0,1)\cup(h^{-1}-1,h^{-1}]}|\eta(t)|>u\}+\sum_{l=1}^{N-1}\mathrm{P}\{\sup_{I_l'}|\eta(t)|>u\}.$$

□

Lemma A.13. Under Assumptions (A1)–(A6), for $u=u_n$ satisfying $2\sqrt{C(K)}N\lambda_n\varphi(u_n)\varphi(0)\to-\log(1-\alpha)$ with $C(K)$ given in (3.5),

$$\lim_{n\to\infty}\mathrm{P}\left\{\sup_{\cup_{l=1}^{N}I_l}|\eta(t)|\le u_n\right\}=1-\alpha.$$

Proof. First, to apply Lemma A.1, we rewrite

$$\mathrm{E}\left[\left(U_u^{\eta}[a_l,b_l]+D_{-u}^{\eta}[a_l,b_l]\right)I_{\{|X(a_l)|\le u\}}\right]=\int_{a_l}^{a_l+2}+\int_{a_l+2}^{b_l}=I_{1l}+I_{2l}.$$

Similar to Lemma A.12, one also can show that as $n\to\infty$,

$$\sup_{1\le l\le N}I_{1l}=\mathcal{O}\{\varphi(u)\}.\tag{A.19}$$

Further, since $r_{a_lt}=r_{1,0}(t,a_l)=0$, $v^2=r_{1,1}(t,t)$ for $\forall t\in[a_l+2,b_l]$, see (A.12), one can simplify (A.18) as

$$I_{2l}=2\varphi(u)\varphi(0)\int_{a_l+2}^{b_l}\sqrt{r_{1,1}(t,t)}dt$$
$$-4\varphi(u)\varphi(0)\{1-\Phi(u)\}\int_{a_l+2}^{b_l}\sqrt{r_{1,1}(t,t)}dt,$$

hence if $2\sqrt{C(K)}N\lambda_n\varphi(u_n)\varphi(0)\to-\log(1-\alpha)$, as $n\to\infty$,

$$\sup_{1\le l\le N}|I_{2l}-2\lambda_n\varphi(u)\varphi(0)\sqrt{C(K)}|=o\{\varphi(u)\lambda_n\}.\tag{A.20}$$

Therefore, (A.19) and (A.20) show that

$$\sup_{1\le l\le N}\Big|\mathrm{E}\left[\left(U_u^{\eta}[a_l,b_l]+D_{-u}^{\eta}[a_l,b_l]\right)I_{\{|X(a_l)|\le u\}}\right]$$
$$-2\lambda_n\varphi(u)\varphi(0)\sqrt{C(K)}\Big|=o\{\varphi(u)\lambda_n\}.\tag{A.21}$$

Now consider the second-order moment and it is easy to verify that

$$\mathrm{E}\left(U_u^{\eta}[a_l,b_l]+D_{-u}^{\eta}[a_l,b_l]\right)^{[2]}$$
$$=2\,\mathrm{E}\,U_u^{\eta}[a_l,b_l]^{[2]}+2\,\mathrm{E}\left(U_u^{\eta}[a_l,b_l]\,D_{-u}^{\eta}[a_l,b_l]\right).$$

By Lemma A.2 and the Hölder inequality

$$\mathrm{E}\,U_u^{\eta}[a_l,b_l]^{[2]}$$
$$=\int_{s,t\in[a_l,b_l]^2}\int_{(0,\infty)^2}|\eta_1'|\,|\eta_2'|\,p_{\eta_t;\eta_s;\eta_t';\eta_s'}(u;u;\eta_1';\eta_2')\,d\eta_1'd\eta_2'dtds$$
$$=\int_{s,t\in[a_l,b_l]^2}\mathrm{E}\{\eta'(t)^+\eta'(s)^+|\eta(t)=\eta(s)=u\}p_{\eta(t),\eta(s)}(u,u)\,dtds$$
$$\le\int_{s,t\in[a_l,b_l]^2}\mathrm{E}^{1/2}[\{\eta'(t)^+\}^2|\eta(t)=\eta(s)=u]$$
$$\times\mathrm{E}^{1/2}[\{\eta'(s)^+\}^2|\eta(t)=\eta(s)=u]p_{\eta(t),\eta(s)}(u,u)\,dtds$$
$$=\int_{2\le|s-t|,s,t\in[a_l,b_l]^2}+\int_{\delta\le|s-t|<2,s,t\in[a_l,b_l]^2}+\int_{|s-t|<\delta,s,t\in[a_l,b_l]^2}$$
$$=I_{1l}+I_{2l}+I_{3l},\tag{A.22}$$

where $p_{\eta(t),\eta(s)}(u,u)=(2\pi\sqrt{1-r^2})^{-1}\exp\{-u^2/(1+r)\}$, see Azaïs and Wschebor (2009) p. 96, Gaussian Rice Formula, and $\delta\in(0,1)$ which does not depend on n, see Lemma A.11.

For I_{1l}, one has $\mathrm{E}[\{\eta'(t)^+\}^2|\eta(t)=\eta(s)=u]\le\mathrm{E}[\{\eta'(t)\}^2|\eta(t)=\eta(s)=u]\le\mathrm{E}^2\{\eta'(t)|\eta(t)=\eta(s)=u\}+\mathrm{var}\{\eta'(t)|\eta(t)=\eta(s)=u\}$ and

$$\mathrm{E}\{\eta'(t)|\eta(t)=\eta(s)=u\}=r_{1,0}(t,s)u/(1+r),\tag{A.23}$$
$$\mathrm{var}\{\eta'(t)|\eta(t)=\eta(s)=u\}=r_{1,1}(t,t)-r_{1,0}^2(t,s)/(1-r^2),$$

see Azaïs and Wschebor (2009) p. 96. If $|t-s|\ge2$, then $r_{st}=r_{1,0}(t,s)=0$ so $\mathrm{E}\{\eta'(t)|\eta(t)=\eta(s)=u\}=0$ and $\mathrm{var}\{\eta'(t)|\eta(t)=\eta(s)=u\}=r_{1,1}(t,t)$. Hence

$$I_{1l}\le\int_{2\le|s-t|,s,t\in[a_l,b_l]^2}\sqrt{r_{1,1}(t,t)}\sqrt{r_{1,1}(s,s)}\frac{1}{2\pi}\exp(-u^2)dtds,\tag{A.24}$$

which implies that

$$\sup_{1\le l\le N}I_{1l}=\mathcal{O}\{\varphi^2(u)\lambda_n^2\}.\tag{A.25}$$

For I_{2l}, similarly,

$$I_{2l}\le\int_{\delta\le|s-t|<2,s,t\in[a_l,b_l]^2}\left\{r_{1,0}^2(t,s)u^2/(1+r)^2+r_{1,1}(t,t)\right\}^{1/2}$$
$$\times\left\{r_{1,0}^2(s,t)u^2/(1+r)^2+r_{1,1}(s,s)\right\}^{1/2}$$
$$\times\frac{1}{2\pi\sqrt{1-r^2}}\exp\{-u^2/(1+r)\}dtds.\tag{A.26}$$

By (A.14), for large n, $\exists c>0$ such that $\sup_{|t-s|\ge\delta>0}(1+r)\le2-c$ and $\inf_{|t-s|\ge\delta>0}|1-r^2|\ge c>0$, so $\exists$ constants $L_1,K_1>0$ such that

$$\sup_{1\le l\le N}I_{2l}\le L_1\varphi\{(1+K_1)u\}\lambda_n.\tag{A.27}$$

One can bound I_{3l} using the inequalities (4.10) and (4.11), Azaïs and Wschebor (2009) p. 97, that is, for $Z\sim\mathrm{N}(\mu,\sigma^2)$, if $\mu>0$, $\mathrm{E}(Z^+)^2\le\mu^2+\sigma^2$ and if $\mu<0$, $\mathrm{E}(Z^+)^2\le(\mu^2+\sigma^2)\{1-\Phi(-\mu/\sigma)\}+\mu\sigma\varphi(\mu/\sigma)$. Since $\eta'(t)$, $\eta'(s)$ conditioning on $\eta(t)=\eta(s)=u$ have a joint Gaussian distribution, see Azaïs and Wschebor (2009) p. 96, we denote

$$\mu_1=\mathrm{E}\{\eta'(s)|\eta(t)=\eta(s)=u\},$$
$$\mu_2=\mathrm{E}\{\eta'(t)|\eta(t)=\eta(s)=u\},\tag{A.28}$$
$$\sigma_1^2=\mathrm{var}\{\eta'(s)|\eta(t)=\eta(s)=u\},$$
$$\sigma_2^2=\mathrm{var}\{\eta'(t)|\eta(t)=\eta(s)=u\}.$$

Next, we claim that while $0<|s-t|<\delta$, μ_1 and μ_2 have opposite signs. In fact, if $0<|s-t|<\delta$, by (A.15), for large n, $r_{1,0}(t,s)\sim(t-s)$ and $r_{1,0}(s,t)\sim(s-t)$ and by (A.14), $\inf_{|t-s|<\delta}(1+r)\ge c>0$, which imply that $\mu_1\mu_2<0$,see (A.23). Further, according to (A.17), (A.23), for large n, $\exists$ constant $L_2>$ such that $\inf_{|t-s|<2,t,s\in[0,h^{-1}]}|\mu_2|\sigma_2^{-1}\ge L_2u$. Without loss of generality, by (A.28), let $\mu_1>0>\mu_2$, then

$$I_{3l}\le\int_{|s-t|<\delta,s,t\in[a_l,b_l]^2}\sqrt{\mu_1^2+\sigma_1^2}\Big[\left(\mu_2^2+\sigma_2^2\right)\{1-\Phi(-\mu_2/\sigma_2)\}$$
$$+\mu_2\sigma_2\varphi(\mu_2/\sigma_2)\Big]^{1/2}\frac{1}{2\pi\sqrt{1-r^2}}\exp\{-u^2/(1+r)\}dtds.$$

It follows from (A.15) and (A.16) that for large enough n, $\exists$ constants $L_3,L_4,L_5,K_2>0$ such that

$$\sup_{1\le l\le N}I_{3l}\le\int_{|s-t|<\delta,s,t\in[a_l,b_l]^2}L_3\sqrt{(s-t)^2u^2+(s-t)^2}$$
$$\times[\{(s-t)^2u^2+(s-t)^2\}\{1-\Phi(L_2u)\}$$
$$-(s-t)^2u\varphi(-L_2u)]^{1/2}|s-t|^{-1}\varphi(u)\,dsdt$$
$$\le L_5\delta\varphi\{(1+K_2)u\}\lambda_n.\tag{A.29}$$

Hence, if $2\sqrt{C(K)}N\lambda_n\varphi(u_n)\varphi(0)\to-\log(1-\alpha)$, as $n\to\infty$, (A.25), (A.27), and (A.29) imply that

$$\sup_{1\le l\le N}\mathrm{E}\,U_u^{\eta}[a_l,b_l]^{[2]}=o\{\varphi(u)\lambda_n\}.$$

Similarly, one has $\mathrm{E}(U_u^{\eta}[a_l,b_l]D_{-u}^{\eta}[a_l,b_l])=o\{\varphi(u)\lambda_n\}$ and then

$$\sup_{1\le l\le N}\mathrm{E}\left(U_u^{\eta}[a_l,b_l]+D_{-u}^{\eta}[a_l,b_l]\right)^{[2]}=o\{\varphi(u)\lambda_n\}.\tag{A.30}$$

In fact, by Lemma A.1, (A.21) and (A.30) show that, as $n \to \infty$,

$$\mathrm{P}\{\sup_{I_l} |\eta(t)| > u\} = 2\sqrt{C(K)}\varphi(u)\varphi(0)\lambda_n + o\{\varphi(u)\lambda_n\}. \quad \text{(A.31)}$$

Finally, since $\mathrm{E}\eta(t)\eta(s) = 0$ while $t \in I_l$, $s \in I_m, l \neq m$, then $\eta(t), \eta(s)$ for $t \in I_l$, $s \in I_m, l \neq m$ are independent Gaussian processes, then $\mathrm{P}\{\sup_{\cup_{l=1}^N I_l} |\eta(t)| \le u\} = \prod_{l=1}^N [1 - \mathrm{P}\{\sup_{I_l} |\eta(t)| > u\}]$ and hence

$$\begin{aligned}
&\mathrm{P}\{\sup_{\cup_{l=1}^N I_l} |\eta(t)| \le u\} \\
&= \exp\left(\sum_{l=1}^N \log\left[1 - \mathrm{P}\{\sup_{I_l} |\eta(t)| > u\}\right]\right) \\
&= \exp\left(\sum_{l=1}^N \log[1 - 2\sqrt{C(K)}\varphi(u)\varphi(0)\lambda_n + o\{\varphi(u)\lambda_n\}]\right) \\
&= \exp[-2N\sqrt{C(K)}\varphi(u)\varphi(0)\lambda_n + o\{N\varphi(u)\lambda_n\}].
\end{aligned}$$

Since $2\sqrt{C(K)}N\lambda_n\varphi(u)\varphi(0) \to -\log(1-\alpha)$ as $n \to \infty$, then it follows from the definitions of N, λ_n, u_n that $\lim_{n\to\infty} \mathrm{P}\{\sup_{\cup_{l=1}^N I_l} |\eta(t)| \le u\} = 1 - \alpha$. □

The quantile $Q_h(\alpha)$ given in (3.5) satisfies $2\sqrt{C(K)}N\lambda_n\varphi\{Q_h(\alpha)\}\varphi(0) \to -\log(1-\alpha)$, as $n \to \infty$, then Lemmas A.12 and A.13 imply that $\lim_{n\to\infty} \mathrm{P}\{\sup_{[0,1]} |\eta(x)| \le Q_h(\alpha)\} = 1 - \alpha$, that is,

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{P}\big[a_h\{\sup_{[0,1]} |\eta(x)| - a_h\} - \log\{\sqrt{C(K)}/(2\pi)\} \\
&\quad \le -\log\{-\log\sqrt{1-\alpha}\}\big] = 1 - \alpha. \quad \text{(A.32)}
\end{aligned}$$

In particular, $\sup_{[0,1]} |\eta(x)| = \mathcal{O}_p(\sqrt{\log n})$.

Lemma A.14. Under Assumptions (A1)–(A6), let $\Delta_{3,n}(x) = \widetilde{\sigma}_n(x)\sigma_n^{-1}(x) - 1, x \in [0,1]$, then $\Delta_{3,n}(x) = \mathcal{U}(h) + \mathcal{U}_{\text{a.s.}}\{\sqrt{\log n/(nh^2)}\}$ and for $\widehat{\varepsilon}(x)$, $\sigma_n^2(x)$ given in (3.4) as $n \to \infty$

$$\begin{aligned}
&\sup_{[0,1]} |\sigma_n^{-1}(x)\{\varepsilon_{N_\mathrm{T}}(x) + \widehat{\xi}(x)\} - \eta(x)| \\
&= \sup_{[0,1]} |\Delta_{3,n}(x)||\eta(x)| = \mathcal{O}_p\{h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)}\}.
\end{aligned}$$

Proof. See the online supplementary materials. □

Proof Of Proposition 3.1. The proof is trivial. □

Proof Of Theorem 3.1. The decomposition (A.2) implies that

$$\sigma_n^{-1}(x)\{\widehat{m}(x) - m(x)\} = \sigma_n^{-1}(x)\{\widetilde{m}(x) - m(x)\} + \sigma_n^{-1}(x)\widetilde{e}(x). \quad \text{(A.33)}$$

As (A.32) implies that $\sup_{[0,1]} |\eta(x)| = \mathcal{O}_p(\sqrt{\log n})$, Lemma A.14 leads to

$$\sup_{[0,1]} \sigma_n^{-1}(x)\left|\varepsilon_{N_\mathrm{T}}(x) + \widehat{\xi}(x)\right| = \mathcal{O}_p(\sqrt{\log n}).$$

and hence by Lemma A.9, $\sup_{[0,1]} \sigma_n^{-1}(x)\left|\widehat{\varepsilon}(x) + \widehat{\xi}(x)\right| = \mathcal{O}_p(\sqrt{\log n})$. Therefore, Lemma A.8 implies that

$$\begin{aligned}
&\sup_{[0,1]} \sigma_n^{-1}(x)\left|\widetilde{e}(x) - \{\widehat{\varepsilon}(x) + \widehat{\xi}(x)\}\right| \\
&= \mathcal{O}_p\{h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)}\}. \quad \text{(A.34)}
\end{aligned}$$

It follows from (A.34), Lemmas A.9 and A.14 that for $t' > 2/5$ (assumption A5),

$$\begin{aligned}
&\sup_{[0,1]} \left|\sigma_n^{-1}(x)|\widetilde{e}(x)| - |\eta(x)|\right| \\
&= \mathcal{O}_p\{h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)} + \sqrt{h}n^{-t'+1/2}\},
\end{aligned}$$

Further, (A.34) and Lemma (A.7) warrants that

$$\begin{aligned}
&\sup_{[0,1]} \left|\sigma_n^{-1}(x)|\widehat{m}(x) - m(x)| - |\eta(x)|\right| \\
&= \mathcal{O}_p\{\sqrt{n}h^{5/2} + h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)} + \sqrt{h}n^{-t'+1/2}\}, \quad \text{(A.35)}
\end{aligned}$$

and therefore

$$\begin{aligned}
&a_h \sup_{x\in[0,1]} \left|\sigma_n^{-1}(x)|\widehat{m}(x) - m(x)| - |\eta(x)|\right| \\
&= \mathcal{O}_p\big[\sqrt{\log h^{-1}}\{\sqrt{n}h^{5/2} + h\sqrt{\log n} + \sqrt{\log^2 n/(nh^2)} \\
&\quad + \sqrt{h}n^{-t'+1/2}\}\big] = o_p(1). \quad \text{(A.36)}
\end{aligned}$$

Finally, by Slutsky's Theorem, (A.32) and (A.36) show that

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{P}\big[a_h\{\sup_{[0,1]} \sigma_n^{-1}(x)|\widehat{m}(x) - m(x)| - a_h\} \\
&\quad - \log\{\sqrt{C(K)}/(2\pi)\} \le -\log\{-\log\sqrt{1-\alpha}\}\big] = 1 - \alpha,
\end{aligned}$$

which is

$$\lim_{n\to\infty} \mathrm{P}\{\sup_{x\in[0,1]} \sigma_n^{-1}(x)|\widehat{m}(x) - m(x)| \le Q_h(\alpha)\} = 1 - \alpha.$$

SUPPLEMENTARY MATERIALS

Supplement to "A Smooth Simultaneous Confidence Corridor for the Mean of Sparse Functional Data": Supplement containing the details of Lemma A.1 and theoretical proofs referenced in the main article.

scbsfda.R: R package containing code to perform SCC and SCI estimations for the mean of sparse functional data.

[Received April 2013. Revised September 2013.]

REFERENCES

Azaïs, J. M., and Wschebor, M. (2009), *Level Sets and Extrema of Random Processes and Fields*, Hoboken, NJ: Wiley. [671]

Bachrach, L., Hastie, T., Wang, W., Narasimhan, B., and Marcus, R. (1999), "Bone Mineral Acquisition in Healthy Asian, Hispanic, Black, and Caucasian Youth: A Longitudinal Study," *Clinical Endocrinology & Metabolism*, 84, 4702–4712. [668]

Bickel, P. J., and Rosenblatt, M. (1973), "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1, 1071–1095. [662]

Bosq, D. (1996), *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, New York: Springer-Verlag. [669]

Cai, T., and Hall, P. (2006), "Prediction in Functional Linear Regression," *The Annals of Statistics*, 34, 2159–2179. [661]

Cao, G., Wang, J., Wang, L., and Todem, D. (2012), "Spline Confidence Bands for Functional Derivatives," *Journal of Statistical Planning and Inference*, 142, 1557–1570. [662,663]

Cao, G., Yang, L., and Todem, D. (2012), "Simultaneous Inference for the Mean Function Based on Dense Functional Data," *Journal of Nonparametric Statistics*, 24, 359–377. [662,663]

Cardot, H., Ferraty, F., and Sarda, P. (2003), "Spline Estimators for the Functional Linear Model," *Statistica Sinica*, 13, 571–591. [661]

Cierco-Ayrolles, C., Croquette, A., and Delmas, C. (2003), "Computing the Distribution of the Maximum of Gaussian Random Processes," *Methodology and Computing in Applied Probability*, 5, 427–438. [662,668,669,670]

Claeskens, G., and Van Keilegom, I. (2003), "Bootstrap Confidence Bands for Regression Curves and Their Derivatives," *The Annals of Statistics*, 31, 1852–1884. [661,662,668]

Csőrgő, M., and Révész, P. (1981), *Strong Approximations in Probability and Statistics*, New York-London: Academic Press. [669]

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall. [663,666]

Fan, J., and Yao, Q. (2005), *Nonlinear Time Series*, New York: Springer-Verlag. [669]

Fan, J., and Zhang, W. Y. (2000), "Simultaneous Confidence Bands and Hypothesis Testing in Varying-Coefficient Models," *Scandinavian Journal of Statistics*, 27, 715–731. [662,668]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Berlin: Springer. [661]

Hall, P. (1993), "On Edgeworth Expansion and Bootstrap Confidence Bands in Nonparametric Regression," *Journal of the Royal Statistical Society*, Series B, 55, 291–304. [666]

James, G. M., Hastie, T., and Sugar, C. (2000), "Principal Component Models for Sparse Functional Data," *Biometrika*, 87, 587–602. [661,668]

Johnston, G. J. (1982), "Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates," *Journal of Multivariate Analysis*, 12, 402–414. [662,668]

Kerkyacharian, G., Nickl, R., and Picard, D. (2012), "Concentration Inequalities and Confidence Bands for Needlet Density Estimators on Compact Homogeneous Manifolds," *Probability Theory and Related Fields*, 153, 363–404. [662]

Kreiss, J. P., and Neumann, M. H. (1998), "Regression-Type Inference in Nonparametric Autoregression," *The Annals of Statistics*, 26, 1570–1613. [661]

Li, Y., and Hsing, T. (2010), "Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data," *The Annals of Statistics*, 38, 3321–3351. [664]

Lucas, P. W., and Diggle, P. J. (1997), "The Use of Longitudinal Data Analysis to Study the Multi-seasonal Growth Responses of Norway and Sitka Spruce to Summer Exposure to Ozone: Implications for the Determination of Critical Levels," *New Phytollogist*, 137, 315–323. [661]

Ma, S., Yang, L., and Carroll, R. J. (2012), "Simultaneous Confidence Band for Sparse Longitudinal Regression," *Statistica Sinica*, 22, 95–122. [662,663,664]

Müller, H. G. (2009), "Functional Modeling of Longitudinal Data," in *Longitudinal Data Analysis, Handbooks of Modern Statistical Methods*, New York: Wiley, pp. 223–252. [661]

Neumann, M. H., and Polzehl, J. (1998), "Simultaneous Bootstrap Confidence Bands in Nonparametric Regression," *Journal of Nonparametric Statistics*, 9, 307–333. [661]

Paul, D., and Peng, J. (2009), "Consistency of Restricted Maximum Likelihood Estimators of Principale Components," *The Annals of Statistics*, 37, 1229–1271. [663]

Piterbarg, V. I. (1996), *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, Providence, RI: American Mathematical Society. [662,669,670]

Sluis, I., Ridder, M., Boot, A., Krenning, P., and Keizer-Schrama, S. (2002), "Reference Data for Bone Density and Body Composition Measured With Dual Energy X Ray Absorptiometry in White Children and Young Adults," *Archives of Disease in Childhood*, 87, 341–347. [668]

Song, Q., and Yang, L. (2009), "Spline Confidence Bands for Variance Function," *Journal of Nonparametric Statistics*, 21, 589–609. [662]

Wang, J., and Yang, L. (2009), "Polynomial Spline Confidence Bands for Regression Curves," *Statistica Sinica*, 19, 325–342. [664]

——— (2007), "Spline-Backfitted Kernel Smoothing of Nonlinear Additive Autoregression Model," *The Annals of Statistics*, 35, 2474–2503. [661]

Wang, N., Carroll, R. J., and Lin, X. (2005), "Efficient Semiparametric Marginal Estimation for Longitudinal/Clustered Data," *Journal of the American Statistical Association*, 100, 147–157. [664]

Wu, W., and Zhao, Z. (2007), "Inference of Trends in Time Series," *Journal of the Royal Statistical Society*, Series B, 69, 391–410. [662,664]

Xia, Y. (1998), "Bias-Corrected Confidence Bands in Nonparametric Regression," *Journal of the Royal Statistical Society*, Series B, 60, 797–811. [662]

Xue, L., and Yang, L. (2006), "Estimation of Semiparametric Additive Coefficient Model," *Journal of Statistical Planning and Inference*, 136, 2506–2534. [661]

Yao, F. (2007), "Asymptotic Distributions of Nonparametric Regression Estimators for Longitudinal or Functional Data," *Journal of Multivariate Analysis*, 98, 40–56. [662]

Yao, F., Müller, H. G., and Wang, J. L. (2005a), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903. [662,663,664]

——— (2005b), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Staistical Association*, 100, 577–590. [662,664]

Zhao, Z., and Wu, W. (2008), "Confidence Bands in Nonparametric Time Series Regression," *The Annals of Statistics*, 36, 1854–1878. [662,664]